AWARD NUMBER: W81XWH-16-1-0390

TITLE: Role of a Novel Family of Short RNAs, tRFs, in Prostate Cancer

PRINCIPAL INVESTIGATOR: MANJARI KIRAN

CONTRACTING ORGANIZATION: RECTOR & VISITORS OF THE UNIVERSITY
CHARLOTTESVILLE VA 22903

REPORT DATE: August 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
August 2017	Annual	1 Aug 2016 – 31 Jul 2017

4. TITLE AND SUBTITLE

Role of a Novel Family of Short RNAs, tRFs, in Prostate Cancer

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0390

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Manjari Kiran

E-Mail: mk9ua@virginia.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

RECTOR & VISITORS OF THE UNIVERSITY

CHARLOTTESVILLE VA
22903-4833

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

tRFs are precisely generated fragments of tRNA which are shown to function by associating to Argonaute proteins. Unlike microRNA, the biogenesis of tRFs is independent of Dicer and Drosha enzymes. Upon meta-analysis of more than 50 small RNA-seq data of different species and cell lines we have previously shown the presence of fragments generated from both 5' and 3' end of the tRNA and called them as tRF-5 and tRF-3, respectively. We are now interested in mining TCGA prostate cancer patients data to identify tRFs and predict potential targets involved in prostate cancer cell migration and proliferation.

15. SUBJECT TERMS

tRF; tRNA-related fragments; Prostate Cancer; Biomarker

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
Unclassified	Unclassified	Unclassified	Unclassified	11	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION:

In 2009, Prof. Dutta and colleagues discovered a tRNA related fragment generated from tRNA trailer sequence involved in cell proliferation in prostate cancer (Lee et al. 2009). Later in 2014, my colleague Dr. Pankaj Kumar reported the presence of tRFs in different human cell lines and different organisms by mining a number of small RNA-Seq data (Kumar et al. 2014). He also showed that tRFs bind to Argonaute proteins and interacts with its targets in a similar way as miRNA by analyzing PAR-CLIP and CLASH data.
In this project, I am elucidating the potential role of tRNA-derived fragments as prostate cancer biomarker. Discovering a new biomarker for prostate cancer is significant because early detection and accurate prognosis is very important to cure the disease without over treating many patients who do not have life-threatening condition. I will first look for the presence of differentially expressed tRFs in prostate cancer patients versus normal and then will predict targets for the top-most differentially expressed tRFs to elucidate its functional role in disease progression.

2. KEYWORDS: tRF; tRNA-related fragments; Prostate Cancer; Biomarker

3. ACCOMPLISHMENTS:

What were the major goals of the project?

Major Task 1: Mining TCGA short RNA raw sequencing data to identify different types of tRNA-derived fragments. (1-6 months)-100% completed
Major Task 2: Predict the targets of tRFs based on sequence similarity. (7-11 months) - 70% completed

What was accomplished under these goals?

The steps involved in TCGA data mining are shown in Figure 1. First, I downloaded all the aligned reads for RNA-Seq performed by miRNA-seq experimental strategy for prostate cancer. There are 551 bam files corresponding to 494 prostate cancer patients. Out of 494 patients, 484 patients are alive and 10 are dead. The paired normal-tumor data is available for 50 patients. There are two patients TCGA-HC-7740 and TCGA-HC-8258 for which three samples are available: 2 corresponding to tumor (01A and 01B) and 1 normal (11A). There is only one patient 'TCGA-V1-A905' with metastatic tumor and the remaining 441 patients have primary tumor. I performed data processing and tRF identification for all the files, from which I am only reporting the results obtained by comparing 50 paired normal-tumor patients. The reads available from TCGA were already trimmed for adapters and mapped against GRCh37 reference genome using BWA-MEM aligners (parameters: samse –n 10) by Marco Marra group from University of British Columbia (Chu et al. 2016). The mapped bam files were then converted to fastq files using bedtools utility with default settings. In order to work with only high quality reads we discarded reads with <30 phred score in 90% of the read length. Now, in the next step, reads were mapped to human tRNA gene to get tRF specific for each patient sample.

The Rigoutsos group in 2014 has previously reported the existence of numerous tRNA full-length lookalike sequence and tRNA incomplete sequences in the human genome (Telonis et al. 2014). So, I mapped reads against full-length reference tRNA set and tRNA-lookalike as well as against incomplete sequence set using MINTmap Version: 1.0 perl script downloaded from https://cm.jefferson.edu/MINTcodes/. MINTmap thus provides separate output file for fragments that are shared between different set and annotate these fragments as ambiguous to alert user for false-positives (Loher, Telonis, and Rigoutsos 2017). MINTmap also reports the abundance of each tRF by calculating reads per million (RPM) which is number of reads mapped to tRF divided by total number of reads in that small RNA-Seq sample per million. I will use this RPM value to compare the tRF expression across different samples.

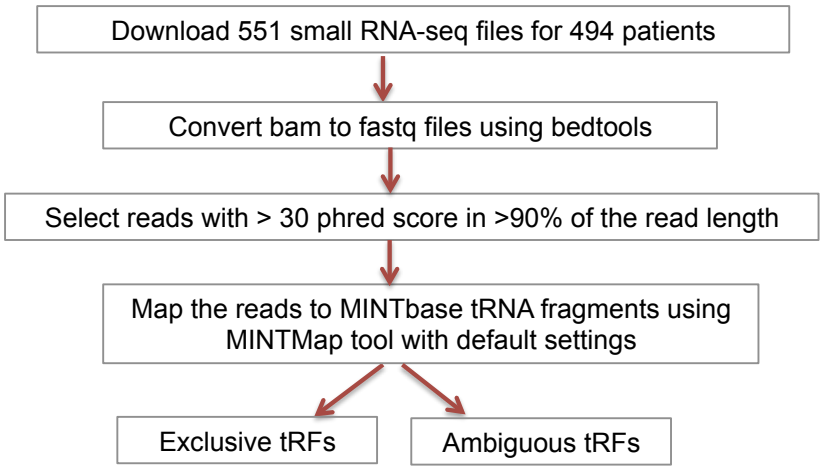


Figure1: Flowchart showing the steps involved in download and processing of data

In order to work with only true positives, I chose a cut-off of 20 RPM and counted combined number of unique tRFs identified by both exclusive and ambiguous method for each patient sample. Around 35 patients have less than 50 tRFs and 25 patients have more than 100 tRFs identified. There are more unique types of tRFs in tumor sample of the patients compared to their normal counterpart (**Figure 2A**).

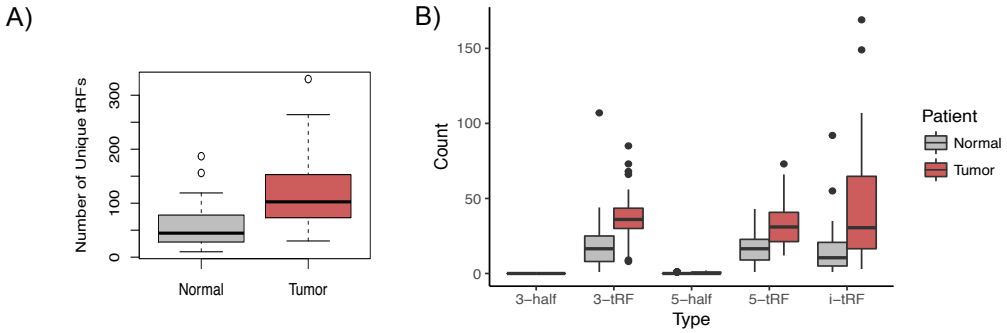


Figure 2: Boxplots showing number of unique tRFs in 50 normal versus 50 tumor patients' samples A) and in different sub-types of tRFs B).

There can be two possibilities explaining this difference: 1) The parent tRNA of the tRFs are more abundantly expressed in tumor than normal. 2) Some unknown factors are more involved in tRF cleavage in tumor samples or more involved in protection in normal samples of the patients. These two possibilities are not mutually exclusive. To check these possibilities, we can compare the abundance of tRFs grouped based of their parent tRNA isoacceptor.

Our group as well as other groups in the field has divided tRNA derived fragments into 5 structural categories.

i) 5-half: longer fragments (>34 nt) that arise from the mature tRNA through cleavage at anticodon of tRNA

ii) 3-half: longer fragments (>34 nt) that are reminder of the mature tRNA following cleavage at anticodon of tRNA

iii) tRF-5/5-tRF: fragments derived after cleavage of mature tRNA at D-loop or the anticodon stem

iv) tRF-3/3-tRF: fragments derived after cleavage of mature tRNA at T-loop or the anticodon stem

v) i-tRF: also known as internal tRFs that can be generated from any other internal sites of tRNA .

I compared the number of distinct types of tRFs in normal and tumor samples of 50 patients. Interestingly, the number of distinct 3-tRF, 5-tRF and i-tRF is significantly higher in tumor than in normal paired samples (P value ~ 2.542e-05) (**Figure 2B**). In contrast, there are no halves identified in either normal or tumor sample. This could be because of running deep-sequencing PCR for only 30 cycles in short RNA-seq library preparation and because of size selection for microRNA sized RNA.

I also noticed higher average expression of tRFs in tumor compared to normal samples (P value = 0.000246) (**Figure 3A**). This again could be because of higher expression or more cleavage of the parent tRNA in tumor than in normal. Among, different structural categories of tRFs, 3-tRFs are the most significantly up-regulated in tumor versus normal (P value ~ 1.387e-05) (**Figure 3B**), which suggests that the cleavage at T-loop is more prominent in tumor samples than normal in prostate cancer patients.

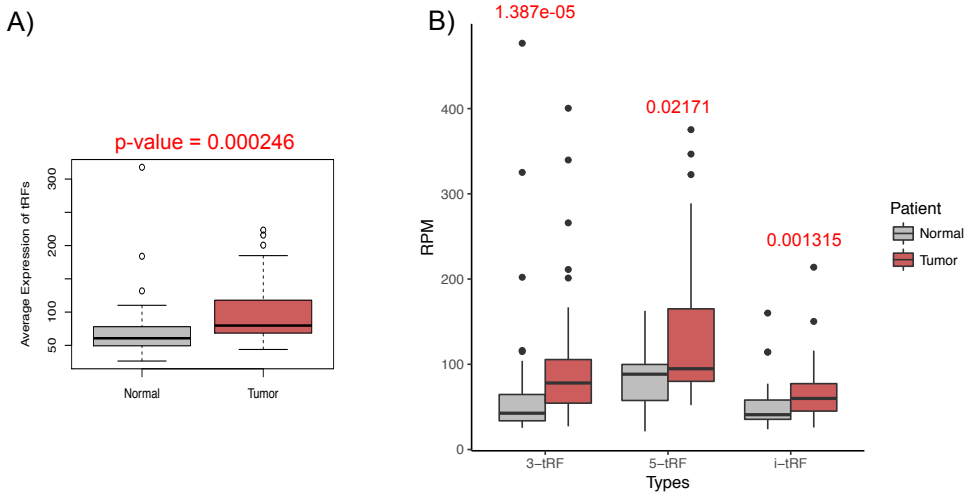
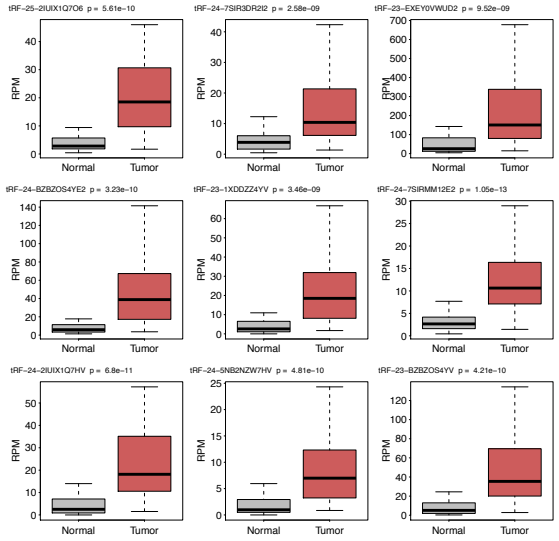


Figure 3: Boxplot showing distribution of average expression of tRFs in 50 normal versus 50 tumor patients' A) and in different sub-types of tRFs B).

My next aim was to find the top most differentially expressed 3-tRFs in tumor versus normal samples. I first filtered out all the 3-tRFs, with mean expression of less than 20 RPM in 50 tumor patients. There were only 63 3-tRFs which met this criteria. Most of these tRFs are 18 bases long that are annotated as tRF-3a in tRFDB. I found 61 3-tRFs which have significantly higher expression in tumors compared to normal. Interestingly, the top-most differentially expressed 3-tRFs are mostly 24 nucleotides long.

A)



B)

tRF Name	tRF sequence	Parent tRNA
tRF-24-7SIRMM12E2	GTTCGATTCCCGGTCAGGGAACCA	GluCTC, GluTTC, PheGAA
tRF-24-2IUIX1Q7HV	CAACTTAACTTGACCGCTCTGACC	ValTAC_MT
tRF-24-BZBZOS4YE2	AACTTAACTTGACCGCTCTGACCA	ValTAC_MT
tRF-23-BZBZOS4YV	AACTTAACTTGACCGCTCTGACC	ValTAC_MT
tRF-24-5NB2NZW7HV	GAGTTAAAGACTTTTTCTCTGACC	ValTAC_MT
tRF-25-2IUIX1Q7O6	CAACTTAACTTGACCGCTCTGACCA	ValTAC_MT
tRF-24-7SIR3DR2I2	GTTCGATTCCCCGACGGGGAGCCA	ProTGG_MT
tRF-23-1XDDZZ4YV	AGTTAAAGACTTTTTCTCTGACC	ValTAC_MT
tRF-23-EXEY0VWUD2	ACTTAACTTGACCGCTCTGACCA	AspGTC

Figure 4: Boxplot showing the distribution of expression level of 9 top-most 3-tRFs obtained by performing Wilcox test which was used to compare mean between normal and tumor prostate cancer patients samples.

Strikingly, more than 70% of 3-tRFs are product of mitochondrial tRNA. 27 and 15 out of 61 differentially expressed 3-tRF are mapping to genomic location of trnaMT_ValTAC_MT_+_1602_1670 and trnaMT_ThrTGT_MT_+_15888_15953, respectively. Further investigation is required to explain this result.

In order to decipher how these fragments actually function, I predicted the targets of top-most differentially expressed 3-tRFs based on sequence complementarity. In our previous study, we have also reported numerous tRF-mRNA chimeras based on CLASH (cross-linking, ligation, and sequencing of hybrids) data analysis, which suggested sequence specific interaction of tRFs with RNAs in the cell in Argonaute containing complexes. With the help of my colleague Dr. Canan Kuscu who is one of the primary experimental persons involved in tRF project in the lab, we mutated the target site on the luciferase reporter three bases at a time. We found that mutations that disrupted the pairing of the target with 5' seed of tRFs failed to repress the target. Mutation M3 and M4 in 2-7 nt region from 5'of tRF disrupted repression the most, presumably by affecting the pairing between tRF and its target. We performed this experiment with multiple other tRFs and found consistent results (**Figure 5**).

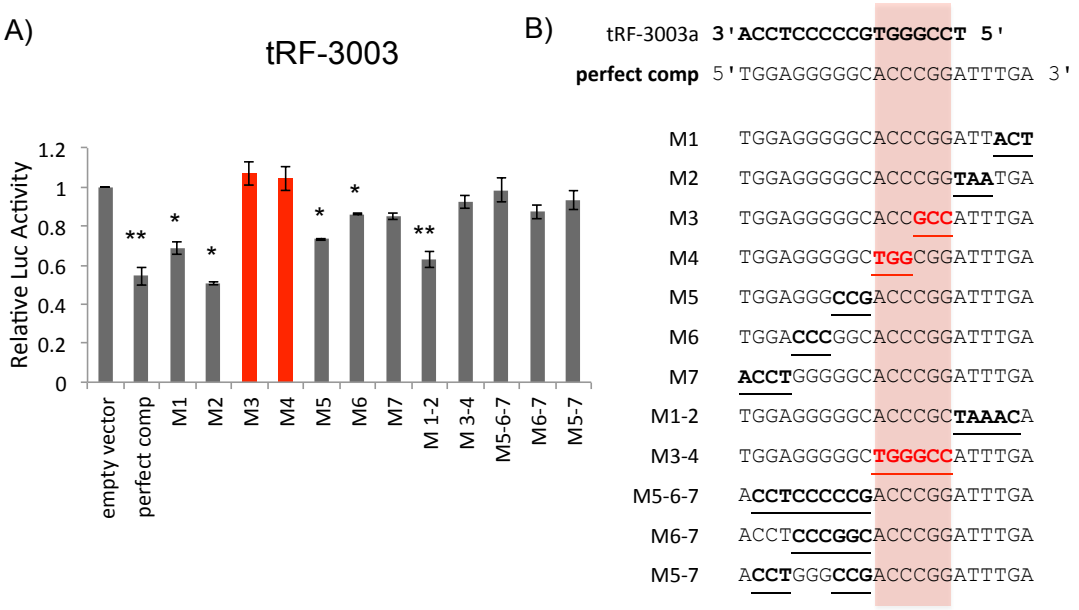


Figure 5: Identification of seed sequence required for target repression by tRFs. A) Luciferase reporter assays with mutant target site at the luciferase reporter upon tRF-3003 overexpression. B) Seed region on tRF-3003 is highlighted in red.

This result suggested that tRFs interact with their targets using their seed sequence similar to miRNA. A script in perl was written to predict targets of the top-most differentially expressed 3-tRFs. The 3'UTR sequence of all RefSeq genes of hg38 genome was downloaded using UCSC Table Browser. In order to remove the bias caused by genes with many isoforms, I considered only the most highly expressed isoform for a gene in Hela cells as identified by 3p-seq by Bartel group in 2014 (Nam et al. 2014). A total of 9294 sequences were examined for the complementarity of various seed sequences. Considering that a tRF interacts with its target using seedmer similar to miRNA, each 3UTR sequence was first scanned for 8mer followed by 7mer-m8, followed by 7mer-A1 and the remaining pool was scanned for 6mer.

In total, I found 2977 targets for tRF-24-2IUIX1Q7HV and 2257 targets for tRF-23-EXEY0VWUD2, the two tRFs identified from previous step as the most differentially expressed in tumor versus normal samples. As expected, due to the difference in seed length and therefore probability to find matching sequence, most of the predicted targets identified belong to 6mer category and least belong to 8mer category. My next aim is to find miRNA and RNA binding proteins as potential targets of these tRFs.

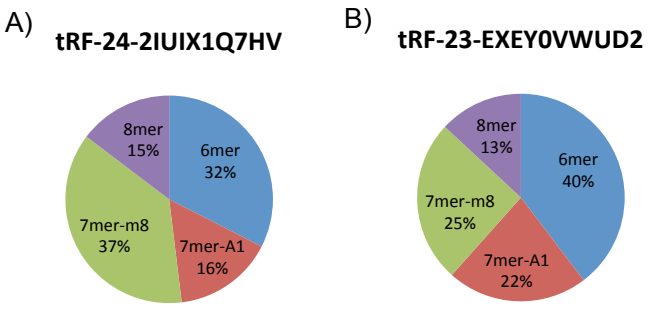


Figure 6: Pie chart showing number of predicted targets for tRF-24-2IUIX1Q7HV A) and tRF-23-EXEY0VWUD2 B)

What opportunities for training and professional development have the project provided?

This project provided me with many opportunities to improve my scientific skills. In the course of 1 year I utilized several computational technique to handle and analyze huge TCGA data. I was also exposed to experimental techniques to answer some of the minor but critical questions asked in the proposal. I have presented my work several times in front of my lab and department, which has helped me in improving my professional communication skill and develop confidence in the project. In conclusion, Dr. Dutta's guidance, productive lab discussions and perfect environment of the lab for pursuing this project are preparing me to be an independent researcher in cancer research.

How were the results disseminated to communities of interest?

"Nothing to Report."

What do you plan to do during the next reporting period to accomplish the goals?

My next major tasks focus on elucidating the prognostic role of tRFs and experimentally validating the functional role of at least five tRFs involved in cell proliferation and migration. For the first task, I will generate a tRF expression profile for all the 494 prostate cancer patients. I will also retrieve clinical information of patients like their vital status: Dead or Alive and Days of last follow-up or disease free status and days to disease free condition. I will use cox-regression to identify tRFs that are associated with the overall survival of the prostate cancer patient. I will then predict targets of these prognostic tRFs and experimentally over-express these tRFs, which have targets involved in cell proliferation, cell migration and invasion to validate their role in prostate cancer.

4. IMPACT:

What was the impact on the development of the principal discipline(s) of the project?

Prostate cancer is the most common cancer in men in United States. The current examination and evaluation procedures are not accurate enough to diagnose prostate cancer progression. This project aims to identify a more specific biomarker for better prognosis of prostate cancer. Small non-coding RNA being short in length, resistant to RNase degradation and longevity in serum can be a promising biomarker. Previous studies have linked many microRNAs to prostate cancer pathogenesis. In 2009, Prof Dutta group identified a tRNA related fragment (tRFs) promoting cell proliferation in prostate cancer. We also know that these tRFs could regulate gene expression in a manner similar to miRNAs. After mining small RNA data available for prostate cancer patient at TCGA, I found many tRFs overexpressed in tumor compared to normal tissue. The results obtained supported the existence of an entirely new group of molecular drivers of prostate cancer. I am also identifying tRFs that can be used for predicting the survival of prostate cancer patient. Such, tRFs can be further studied and could serve as biomarker for early cancer detection or prognosis. The preliminary data obtained in this part of project will help me in designing the future experiments in a more definitive way.

What was the impact on other disciplines?

"Nothing to Report."

What was the impact on technology transfer?

"Nothing to Report."

What was the impact on society beyond science and technology?

"Nothing to Report."

5. CHANGES/PROBLEMS: "Nothing to report"

6. PRODUCTS:
Journal publication:
Kuscu C[1], Kumar P[1], **Kiran M[1]**, Z Su[1], A Malik[1], A Dutta[1]. Global Gene Repression By Dicer-Independent tRNA Fragments. bioRxiv, 143974 (under review)
1Department of Biochemistry and Molecular Genetics, University of Virginia School of Medicine, Charlottesville, VA, USA

Books or other non-periodical, one-time publications. "Nothing to report"

Other publications, conference papers, and presentations:
Dutta A, Kumar P, **Kiran M**, Kuscu C. Transfer RNA Fragments (tRFs): a Novel Class

of Non-micro Short RNAs that Uses Ago1, 3 and 4 to Repress Specific Target RNAs Through 5′ Seed Sequences.The FASEB Journal 30 (1 Supplement), 1054.5-1054.5 (This abstract is from the Experimental Biology 2016 Meeting)

Website(s) or other Internet site(s) "Nothing to report"

Technologies or techniques "Nothing to report"

Inventions, patent applications, and/or licenses "Nothing to report"

Other Products "Nothing to report"

7. PARTICIPANTS & OTHER COLLABORATING ORGANIZATIONS

What individuals have worked on the project?

Name: Manjari Kiran "no change"

Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?

"Nothing to report."

What other organizations were involved as partners?

"Nothing to report."

8. SPECIAL REPORTING REQUIREMENTS None

9. APPENDICES: